



# My Progress on the Final Project

When I first began the project, I knew that I wanted to use a dataset that I am considering to use for my capstone on. It is a dataset of tweets done by suspected ISIS sympathizers.

When I began the process of sketching my ideas, I initially planned on creating a dashboard that would allow me to look at specific tweets and users and gain more insight on them.

Here are some of my earliest sketches:

IMG\_0713

IMG\_0714

IMG\_0715

I knew that this was lacking and needed help so I journeyed to office hours where Professor Field highlighted a potential issue that I knew that I had but was struggling to articulate.

In essence, I was viewing the data at a micro level and not a macro level. In other words, I wasn't looking at the dataset as a whole. This lead me to take a deeper look at the data and determine if I could make any patterns out of it. To do this, I knew that I wanted to generate more information. I decided to initially generate 6 new things.

- `description-sentiment`
- Each entry was given a profile description and I reported the sentiment score.
- `description- subjectivity`
- For the profile description, report subjectivity through pertained model.

*Same as above but for the actual tweet content*

- `tweet-sentiment`
- `tweet- subjectivity`

Next, I used a python package called spacy, to perform entity recognition in the tweets to see if I could pull out any mentions of companies, people, or countries.

*Here's an example from Spacy's Documentation.*

image-20210420165327136

- `People in tweet`
- `Organizations in tweet`
- `"NORP" (Nationalities or religious or political groups.) in tweet`

Now, since I have all this new information, I can recreate the dashboard at the micro level.

The plan is to take features built in the dataset and compare them with the new information that I have generated.

For instance, one thing that I could do is have a location page. (*Location is built into the dataset*) On this page, I could display which people are mentioned the most frequently in a location. I can also ask if certain locations have a more negative text. For instance, I assume that America is talked about extremely negatively, is this true?

## Next Steps

I'm still working on narrowing down the scope and determine which questions that I want to ask. Ultimately, I now have the ability to easily query the data in certain ways and am looking forward to laying everything-out together.

## Appendix

*Here's some of the quick analysis that I have done.*

- Mentions of Nationalities image-20210420170320268
- Mentions of Organizations image-20210420170344413
- Mentions of People image-20210420170405520

*Still need to do some cleaning here- appears to be stumped by camel case and some locations.*